# Framework for Identification and Prevention of Direct and Indirect Discrimination using Data mining

Mr. A.I.Sheikh
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*sheikh.ansar007@gmail.com*

Ms.Monika Kohale
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*monikakohale2@gmail.com*

Ms.AishwaryaWadurkar
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*aishwaryawadurkar07@gmail.com*

Ms. Ashanka Bute
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*ashankabute@gmail.com*

Ms. Dhanashri Baramkar
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*dhanashri23baramkar@gmail.com*

Ms. Aishwarya Ainchwar
Department of Computer Science and Engineering
Smt. Rajshree Mulak College of Engineering for Women, RTMNU
Nagpur, India
*aishwaryaainchwar123@gmail.com*

***Abstract*****—**Extraction of useful and important information from huge collection of data is known as data mining. Negative social perception about data mining is also there, among which potential privacy invasion and potential discrimination are there. Discrimination involves unequally or unfairly treating people on the basis of their belongings to a specific group. Automated data collection and data mining techniques like classification rule mining have made easier to make automated decisions, like loan granting/denial, insurance premium computation, etc. If the training data sets are biased in what regards discriminatory (sensitive) attributes like age, gender, race, religion, etc., discriminatory decisions may ensue. For this reason, antidiscrimination techniques including discrimination discovery, identification and prevention have been introduced in data mining. Discrimination may of two types, either direct or indirect. Direct discrimination is the one where decisions are taken on basis of sensitive attributes. Indirect discrimination is the one where decisions are made based on non-sensitive attributes which are strongly correlated with biased sensitive ones. In this paper, we are dealing with discrimination prevention in data mining and propose new methods applicable for direct or indirect discrimination prevention individually or both at the same time. We discuss how to clean training data sets and transformed data sets in such a way that direct and/or indirect discriminatory decision rules are converted to legitimate (non-discriminatory) classification rules. We also propose new measures and metrics to analyse the utility of the proposed approaches and we compare these approaches.

***Keywords****—Discrimination, Sensitive, Non-sensitive, Anti-discrimination, Dataset.*

*****

## I. Introduction

In sociology, discrimination is the unfair treatment of a person based on their membership in a certain group. It involves denying to members of one group opportunities that are applicable to other groups. There is a list of anti-discriminatory laws designed to prevent discrimination on the basis of a number of attributes (e.g., race, religion, gender, nationality, caste, disability, and age) in various sectors (e.g., employment and training, access to public services, credit and insurance, etc.). For example, the European Union implements the principle of equivalent treatment between men and women in the access to and supply of goods and services in or in matters of employment and occupation. Although some laws and policies are also there toprevent discrimination, many of them fails at certain extent. Certain technologies can add proactivity to legislation by contributing discrimination discovery and prevention methods.

Services in the information society allow for automatic and routine collection of huge amount of data. Those data are often used to train association/classification rules in view of making automated decisions, like loan granting/denial, insurance premium, personnel selection, etc. At first sight, automated decisions may give a sense of equity, classification rules do not guide themselves by personal. However, onecan realize that classification rules are actually learned by the system (e.g., loangranting) from thetraining dataset. If the training data are against a particular community (e.g., foreigners), the learned model may show a discriminatory unfair behavior. In other words, the system may conclude that just being foreigner is a valid reason for loan refusal. Discovering such potential biases and eliminating them from the training data without affecting their decision- making utility is therefore highly desirable. One must avoid data mining from being itself a source of discrimination, due to data mining tasks generating discriminatory models from biased data sets as part of the automated decision making.

Discrimination can be of two types direct or indirect. Direct discrimination consists of rules and procedures that absolutely mention minority or disadvantaged groups which are based on sensitive discriminatory attributes related to a particular group. Indirect discrimination consists of rules and procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could produce discriminatory decisions.

## II. PROBLEM DEFINITION & OVERVIEW

Following are the approaches to prevent -

1. Direct Discrimination Prevention approach.
2. Indirect Discrimination Prevention approach.

Direct Discrimination Prevention Approach:-In this module decisions are made based on sensitive attributes.

Indirect Discrimination Prevention approach:- In Indirect discrimination decisions are made based on non-sensitive attributes which are strongly related with sensitive ones.

### *Problem Definition*

For large amount of data automatic and routine collection is allowed by the information society. In making automated decisions, this data is used to train association and classification rules. Automated decisions can be like loan granting or rejecting, insurance premium, personnel selection, etc. A sense of equity is given by automating decisions at first sight, classification rules can't guide themselves by personal preferences. But, it is realized that the classification rules are originally learned by the system from the training dataset. The learned model can show a discriminatory unfair behavior if the training data are biased for or against a particular community or group.

That is nothing but system may conclude that reason for loan rejection is just for being foreigner. It is highly desirable that the discovering potential biases and removing them from the training dataset without harming their decision making utility. As the data mining tasks generating discriminatory models from biased data sets as part of the automated decision making, we should avoid data mining from becoming itself a source of discrimination. It is experimentally proved in that the data mining can be both a source of discrimination and a means for discovering discrimination.

## III. LITERATURE SURVEY

*[1]Sara Hajian and Josep Domingo-Ferrer ,A Methodology for Direct and Indirect Discrimination Prevention in Data Mining,IEEE Transactions On Knowledge And Data Engineering, VOL.25, NO. 7, JULY 2013*

The above paper states, Data mining is a very useful technology for extracting useful information hidden in large collections of data. There are, however, negative social perceptions about data mining, among which potential privacy invasion and potential discrimination.

*[2]Supriya M.Manglekar, V. K. Bhusari ,Data mining for Discrimination Prevention ,International Journal of Advanced Research in Computer Science and Software Engineering ,VOLUME 4, ISSUE 11, NOVEMBER 2014*

For privacy, discrimination is an important issue when considering the legal and ethical aspects of data mining. It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, caste, age and so on, especially when those attributes are used for making decisions about them like giving them a job, loan, insurance, etc.

[3]*K. Madhavi, Krishna Naik Mudavath,Prevention Methods for Discrimination in Data Mining,International Journal of Science and Research (IJSR),VOLUME 3, ISSUE 11, NOVEMBER 2014*

Discrimination is an important issue when considering as law and professional aspects of data mining. Data mining is an increasing crucial technology for extracting useful information hidden in huge collections of data and negative social perceptions in data mining. This is providing along with privacy and security of the information and potential discrimination.

## IV. PROPOSED SYSTEM

We propose new utility measures for prevention of both direct and indirect discrimination.

Based on the proposed measures, we will present extensive experimental results for two well-known data sets and compare the different possible techniques for direct or indirect discrimination prevention to find out which methods could be more successful in terms of low information loss and high discrimination removal. This approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination.

Our Proposed data transformation methods rule protection and rule generalization are based on measures for both direct and indirect discrimination and can deal with different discriminatory items.

We present a unified approach to direct and indirect discrimination anticipation, with finalized algorithms and all possible data transformation methods based on rule protection or rule generalization that could be applied for direct or indirect discrimination prevention.

## V. METHODOLOGY

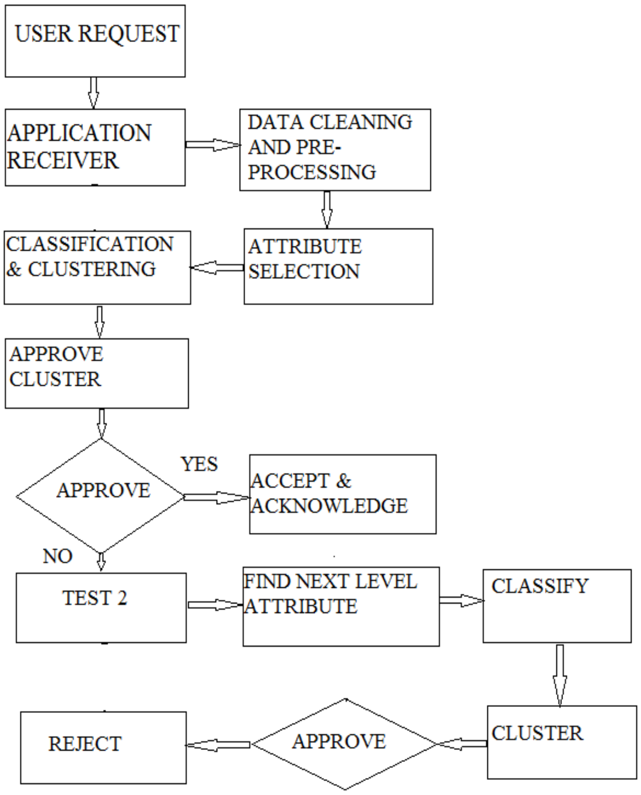

Fig.1.Data flow diagram

Following are the modules to be used in the project:

1. Data Collection**:** A dataset is a collection of data objects (records) and their attributes.
2. Data cleaning and pre-processing: Data is cleansed through processes such as filling missing values, smoothing noisy data or resolving inconsistency in data. Data pre-processing is data mining technique that involves transforming raw data into an understandable format.
3. Feature selection: selection of appropriate rules which are defined.
4. Data classification: It is the process of organizing data into categories for its most effective and efficient use.
5. Graph generation: Creating the graph according to the selected rules.
6. Algorithm testing: Testing the data according to Naïve Bayes algorithm.

*Algorithm:*

Naïve Bayes Algorithm:- It is a classification techniques based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a naïve Bayes classifiers assume that the presence of a particular feature in a class is unrelated to the existence of any other feature .this model is easy to build and particularly useful for very large data sets. Along with simplicity,Naïve Bayes is known to outperform even highly sophisticated classification method.

In this section, we describe our algorithms based on the direct and indirect discrimination prevention methods

*Direct and Indirect Discrimination Prevention Algorithms:*

This Algorithm details our proposed data transformation method for simultaneous direct and indirect discrimination prevention. The algorithm starts with redlining rules. From each redlining rule (r: X! C), more than one indirectα-discriminatory rule (r0: A; B! C) might be generated because of two reasons: 1) existence of different ways to cluster the items in X into a context item set B and a non-discriminatory item set D correlated to some discriminatory item set A; and 2) existence of more than one item in DIs. Hence, as shown in Algorithm (Step 5), given a redlining rule r, proper data transformation should be conducted for all indirect α -discriminatory rules r': (A ⊆ Dis), (B⊆ X) ->C ensuing from r.

*Algorithm*:Direct and Indirect Discrimination Prevention

```
1: Inputs: DB, FR, RR, MR,α, DIs
2: Output: DB' (transformed data set)
3: for each r: X -> C ∈RR, where D, B ⊆X do
4: γ = con f(r)
5: for each r': (A ⊆ Dis), (B ⊆ X) ->C∈ RRdo
6: β2 =conf (rb2: X -> A)
7: Δ1 =supp (r2: X -> A)
8: δ = con f (B ->C)
9: Δ2 = supp (B -> A)
10: β1 = Δ1/Δ2 //con f (rb1: A, B -> D)
11: Find DBc: all records in DB that completely support ¬ A; B;¬D ->¬ C
12: Steps 6-9 Algorithm 1
13: if r'∈MRthen
14: while (δ<=β2 (β2+γ-1/β2-α) and (δ<=con f (r')/α) do
 15: Select first record dbc in DBc
 16: Modify the class item of dbc from¬ C to C in DB
 17: Recompute δ= con f (B -> C)
18: end while
19: else
20: while δ<=β2 (β2+γ-1) /β2-αdo
21: Steps 15-17 Algorithm 4
22: end while
23: end if
24: end for
25: end for
 26: for each r`: (A, B -> C)∈ MR \ RR do
 27: δ= con f (B -> C)
28: Find DBc: all records in DB that completely support: A, B ->¬ C
 29: Step 12
 30: while (δ <= con f (r')/α) do
```

```
31: Steps 15-17 Algorithm 4
32: end while
33: end for
 34: Output: DB` = DB
```

If some rules can be drawn fromDB as both direct and indirect α -discriminatory rules, it means that there is overlap between MR and RR; in such case, data transformation is performed until both the direct and the indirect rule protection requirements are fulfilled (Steps 13- 18). This is possible because, the same data transformation method (Method 2 consisting of changing the class item) can provide both DRP and IRP. However, if there is no overlap between MR and RR, the data transformation is performed according to Method 2 for IRP, until the indirect discrimination prevention requirement is satisfied (Steps 19-23) for each indirect α-discriminatory rule ensuing from each redlining rule inRR; this can be done without any negative impact on direct discrimination prevention, as justified. Then, for each direct α-discriminatory rule r'∈ MR\RR (that is only directly extracted from DB), data transformation for satisfying the direct discrimination prevention requirement is performed (Steps 26-33), based on Method 2 for DRP; this can be done without any negative impact on indirect discrimination prevention, as justified. Performing rule protection or generalization for each rule in MR by each of Algorithms 1-4 has no adverse effect on protection for other rules (i.e., rule protection at Step i+x to make r' protective cannot turn into discriminatory a rule r made protective at Step i) because of the two following reasons: the kind of data transformation for each rule is the same (change the discriminatory item set or the class item of records) and there are no two α-discriminatory rules r and r' in MR such that r =r'.

## VI. CONCLUSION

The purpose of this projectis to develop a new pre- processing discrimination prevention methodology containing different data transformation methods that can avoid direct discrimination, indirect discrimination or both of themat thesametime.To achieve thisobjective,thefirst step is to measure discrimination and identify the categories and groups of individuals that have been directly or indirectly discriminated in the decision-making processes; the second step is to transform data in the appropriate way to remove all those discriminatory biases. Finally, discrimination-free data can be formed from the altered data set without seriously damaging data quality. The experimental results reported demonstrate that the proposed methods are quite successful in both goals of removing discrimination and preserving data quality.

## REFERENCES

[1] D. Pedreschi, S.Ruggieri, F. Turini, "Discrimination-Aware Data Mining", Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining , pp. 560-568, 2008.

[2] D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discrimination in Socially-Sensitive Decision Records", Proc. Ninth SIAM Data Mining Conf. (SDM '09), pp. 581-592, 2009.

[3] European Commission, "EU Directive2006/54/EC on Anti-Discrimination", http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri= OJ:L:2006:204:0023:0036:en:PDF, 2006.

[4] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling", Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[5] F. Kamiran, T. Calders, and M. Pechenizkiy, "Discrimination Aware Decision Tree Learning," Proc. IEEE Int'l Conf. Data Mining (ICDM '10), pp. 869-874, 2010.

[6] Kamiran and T. Calders, "Classification without Discrimination", Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009.

[7] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proc 20th Int'l Conf. Very Large Databases, pp.487-499, 1994.

[8] S. Hajian, J. Domingo-Ferrer, and A. Martinez-Balleste, "Rule Protection for Indirect Discrimination Prevention in Data Mining," Proc. Eighth Int'l Conf. Modeling Decisions for Artificial Intelligence (MDAI '11), pp. 211-222, 2011.

[9] Sara Hajian and Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", IEEE Transaction on Knowledge and Data Engineering, Vol.25, No 7, July 2013.

[10] S. Ruggieri, D. Pedreschi and F. Turini, "Data Mining for Discrimination Discovery", ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

[11] T. Calders and S. Verwer, "Tree Naïve Bayes Approaches for Discrimination Free Classification",